Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2010;39:1567–1581
doi:10.1093/ije/dyq136

Efficacy of experimental treatments compared with standard treatments in non-inferiority trials: a meta-analysis of randomized controlled trials

Darius Soonawala,[1,2]* Rutger A Middelburg,[1] Matthias Egger,[3] Jan P Vandenbroucke[1] and Olaf M Dekkers[1,4]

[1]Department of Clinical Epidemiology, Leiden University Medical Centre, RC Leiden, The Netherlands, [2]Department of Infectious Diseases, Leiden University Medical Centre, RC Leiden, The Netherlands, [3]Institute of Social and Preventive Medicine, University of Bern, Berne, Switzerland and [4]Department of Endocrinology and Metabolic Diseases, Leiden University Medical Centre, RC Leiden, The Netherlands

*Corresponding author. Department of Infectious Diseases C5-P, Leiden University Medical Centre, Albinusdreef 2, 2300 RC Leiden, The Netherlands. E-mail: d.soonawala@lumc.nl

Accepted 6 July 2010

Background There is concern that non-inferiority trials might be deliberately designed to conceal that a new treatment is less effective than a standard treatment. In order to test this hypothesis we performed a meta-analysis of non-inferiority trials to assess the average effect of experimental treatments compared with standard treatments.

Methods One hundred and seventy non-inferiority treatment trials published in 121 core clinical journals were included. The trials were identified through a search of PubMed (1991 to 20 February 2009). Combined relative risk (RR) from meta-analysis comparing experimental with standard treatments was the main outcome measure.

Results The 170 trials contributed a total of 175 independent comparisons of experimental with standard treatments. The combined RR for all 175 comparisons was 0.994 [95% confidence interval (CI) 0.978–1.010] using a random-effects model and 1.002 (95% CI 0.996–1.008) using a fixed-effects model. Of the 175 comparisons, experimental treatment was considered to be non-inferior in 130 (74%). The combined RR for these 130 comparisons was 0.995 (95% CI 0.983–1.006) and the point estimate favoured the experimental treatment in 58% ($n = 76$) and standard treatment in 42% ($n = 54$). The median non-inferiority margin (RR) pre-specified by trialists was 1.31 [inter-quartile range (IQR) 1.18–1.59].

Conclusion In this meta-analysis of non-inferiority trials the average RR comparing experimental with standard treatments was close to 1. The experimental treatments that gain a verdict of non-inferiority in published trials do not appear to be systematically less effective than the standard treatments. Importantly, publication bias and bias in the design and reporting of the studies cannot be ruled out and may have skewed the study results in favour of the

experimental treatments. Further studies are required to examine the importance of such bias.

Keywords Non-inferiority, meta-analysis, systematic review

Introduction

Non-inferiority trials are increasingly published in the medical literature, increasingly used in drug licensing and have at the same time come under increased scrutiny and criticism, up to the allegation that they are unethical.[1–13] A verdict of 'non-inferiority' leaves readers with the impression that a new experimental treatment is as good as an established standard treatment and that the two can be used interchangeably. However, in such trials, non-inferiority is statistically accepted whenever an experimental treatment is unlikely to be worse than an established treatment by more than a pre-specified amount, the so-called non-inferiority margin. If a relatively wide margin is chosen, new treatments that are actually less beneficial might wrongly be considered as equally effective. This may lead to acceptance and use of new therapies that are actually less effective in a clinically relevant way.[10,12]

There is concern that non-inferiority trials might be deliberately designed to conceal that a new treatment is somewhat less effective than a standard treatment.[10,12] Systematic use of too-large non-inferiority margins or systematic biases of design, conduct or reporting of non-inferiority trials may skew results in favour of new treatments.[14–19] In this meta-analysis we examined one type of systematic bias. If trialists systematically compare slightly less effective new treatments with standard treatments, the combined results from a meta-analysis of many trials in which experimental treatments gain a verdict of non-inferiority, would be expected to favour the standard treatment. In order to test this hypothesis, we performed a meta-analysis of non-inferiority trials published in clinical journals and assessed the average effect of experimental treatments compared with standard treatments. Importantly, the combined estimate from the meta-analysis will not be influenced by the choice of the non-inferiority margins.

Methods

Eligibility and search strategy

We searched for non-inferiority trials on 20 February 2009 using PubMed (National Library of Medicine) with the text words 'noninferiority' or 'non inferiority' or 'equivalence' combined with the text words 'clinical trial' or 'trial' or 'trials' or 'study' or 'studies', limiting the search to publications from 1991 onwards. The initial search was restricted to six general medicine journals (*Annals of Internal Medicine, BMJ, JAMA, Lancet, New England Journal of Medicine and PLoS Medicine*). In a second step the search was extended to include the other 115 journals included in PubMed's selection of core clinical journals (see http://www.nlm.nih.gov/bsd/aim.html for a list of these journals). Although equivalence trials (trials that specify both a lower and an upper equivalence margin)[20,21] were not eligible for inclusion, we included the term 'equivalence' in our search strategy in order not to miss non-inferiority trials that had been described as equivalence trials.

Study selection

All two-arm parallel group non-inferiority trials of an experimental treatment compared with standard therapy were included, independent of the intervention examined in the trial. Articles that were published electronically ahead of print were also considered.

Data extraction

The following information was extracted independently by two investigators (D.S. and R.M.): year of publication, journal, subject area (cardiovascular medicine, infectious diseases, obstetrics and gynaecology, rheumatology, surgery or other), primary endpoint, non-inferiority margin for primary endpoint, expected incidence of the primary endpoint in the standard arm and the point estimate for the comparison of experimental with standard treatment. The primary endpoint was classified into three categories: (i) mortality alone or as part of a combined endpoint; (ii) clinical disease; and (iii) surrogate endpoint (imaging or laboratory test). Disagreements were resolved in consultation with a third investigator (O.D.). If trials presented more than one primary endpoint, the endpoint for which the sample size had been calculated was used. If it was unclear for which endpoint sample size calculations were done, or if no such calculations were reported, one of the primary endpoints was randomly selected. In trials that included several non-inferiority comparisons using the same standard treatment, e.g. when testing two dosages of an experimental therapy, one comparison was selected at random and included in the analysis. If a study reported both intention-to-treat- and per-protocol analyses, the result used by the author to determine whether the intervention was non-inferior was extracted. If this was not clear, the per-protocol results were used. The funding source was independently classified by two investigators (D.S. and O.D.) as industry, public or mixed funding. The provision of

study drugs by industry was considered as a source of industry funding.

Data synthesis and analysis

The confidence intervals (CIs) and non-inferiority margins reported by the investigators were used to classify the results as superior, non-inferior, inconclusive or inferior according to the definitions given in the extension of the Consolidated Standards of Reporting Trials (CONSORT) statement to non-inferiority trials.[9] Superiority was assumed if the experimental treatment was significantly ($P < 0.05$) more efficacious than the standard treatment. Non-inferiority was assumed if the 95% CI did not include the non-inferiority margin. Results were classified as inconclusive if the 95% CI included the non-inferiority margin. Treatments were assumed to be inferior if the entire 95% CI was significantly worse than the non-inferiority margin.

Results of comparisons were expressed as ratio measures, which we call relative risks (RRs) throughout this article. If the trial reported risk ratios or hazard ratios (HRs) or odds ratios (ORs) from statistical models these were used in the analyses. For trials that reported risk differences, we calculated the risk ratio. For studies reporting continuous endpoints (e.g. blood pressure), results were converted to ORs using the method described by Hasselblad and Hedges,[22] and the OR was then used in further calculations. This method assumes logistic distributions with equal variances in the two treatment groups. Under this assumption the natural logarithm of the OR equals a constant multiplied by the standardized difference between means. If needed, the inverse of the RR was calculated, so that ratios >1 consistently favoured the standard treatment and ratios <1 favoured the experimental treatment. The RRs from individual studies were combined using random-effects models. In addition to combining RRs for all studies, we performed stratified meta-analyses according to whether results were interpreted as inferior, non-inferior or superior, by type of effect measure, by type of endpoint, by source of funding, by journal and according to whether the judgement of the result was based on an intention-to-treat analysis or not. In a random-effects meta-regression model we analysed the influence of the source of funding.

Pre-specified non-inferiority margins were also expressed as RRs. Margins that were reported as a difference in incidence were converted to RR by dividing the expected incidence of the primary endpoint in the standard treatment arm plus (for morbidity or mortality) or minus (for beneficial endpoints) the pre-specified margin by the expected incidence of the primary endpoint in the standard treatment arm. For example, if the expected mortality rate in the standard treatment arm was 10% and the pre-specified margin was set at 2%, the margin converted to an RR of 1.2 [(10 + 2)/10]. Margins could not be expressed as RR for studies that did not report the expected incidence, or studies reporting continuous endpoints. We examined the median and the distribution of non-inferiority margins and examined whether margins differed across the subgroups of trials mentioned above.

We compared the observed incidence of the primary endpoint in the group that received standard treatment with the expected incidence, as specified by the trialists. The result was expressed as a ratio. If needed, the inverse of this ratio was calculated, so that ratios <1 consistently indicated that the standard treatment performed better than was expected at the design stage of the trial, and ratios >1 indicated that the standard treatment performed worse than was expected. This ratio could not be calculated for studies that did not report separately the expected, or the observed incidence of the primary endpoint, or studies reporting continuous endpoints.

Statistical analyses were done in Comprehensive Meta-Analysis (version 2.0, Biostat, Englewood, NJ, USA) and Stata (version 10.0, Stata Corporation, College Station, TX, USA).

Results

Literature search and study characteristics

We identified 532 potentially eligible articles and excluded 362 studies for the reasons shown in Figure 1. A total of 170 studies, which were published in 43 different journals, were included (see Appendix Tables 1, 2 and 3 for bibliographic details[45–197] available as supplementary data at *IJE* online). Five articles reported the results for two separate comparisons. In total, 175 comparisons were therefore included in the analyses. The oldest non-inferiority study in our selection dates from 1993.[23] Seventy-eight percent of included studies date from 2004 onwards, reflecting an increase in non-inferiority trials in the past 5 years.

The characteristics of the 170 non-inferiority trials are described in Table 1. Of the general medical journals, the *New England Journal of Medicine* published the largest number of non-inferiority trials; our search found no non-inferiority trials in *PLoS Medicine*. Most trials were from cardiovascular medicine ($n = 47$; 28%) or infectious disease ($n = 43$; 25%). Other fields were obstetrics and gynaecology (7%), oncology (6%), rheumatology (6%), surgery (5%), psychiatry (3%), general medicine (3%), pulmonary medicine (2%), gastroenterology (2%), anaesthesiology (1%), intensive care medicine (1%) and neurology (1%). The majority reported risk differences ($n = 106$; 61%); 36 studies reported continuous endpoints, which were converted to ORs for the present analysis. Figure 2 shows the type of endpoint, number of participants, point estimate, CI and pre-specified margin

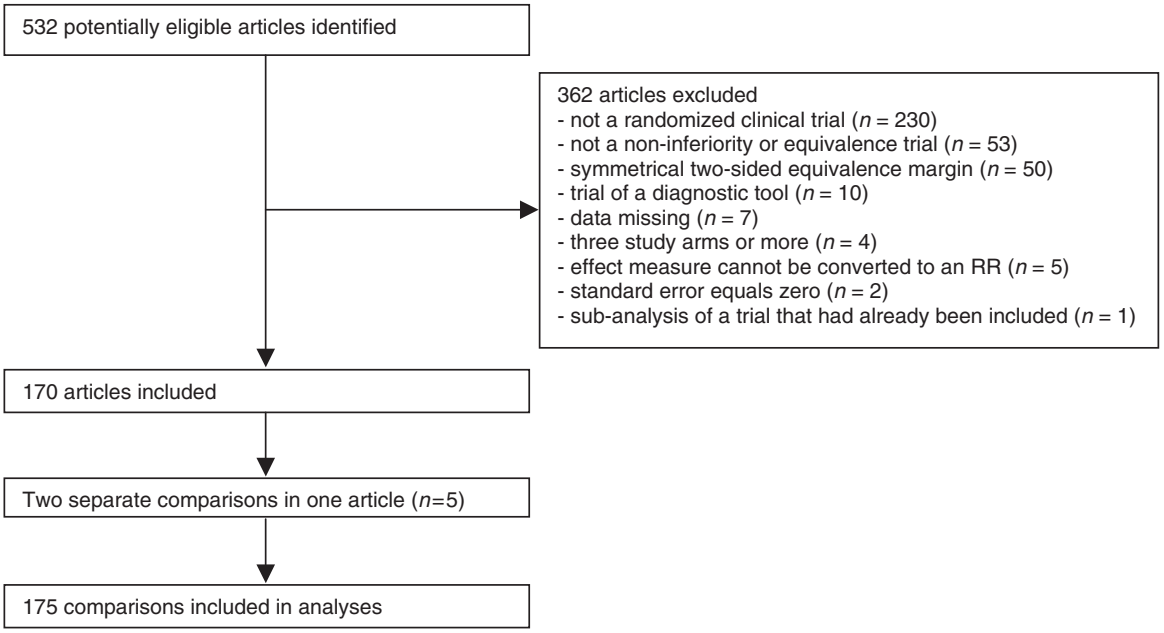

Figure 1 Summary of the search strategy and study selection

for 33 comparisons that were reported on the RR scale.

For 130 comparisons (74%), we considered the experimental treatment to be non-inferior according to the published criteria.[9] Of note, in 6 of these 130 comparisons the authors deemed the experimental treatment to be clinically inferior based on a secondary endpoint. For 27 comparisons (15%) results were inconclusive, for 15 comparisons (9%) superior and for 3 comparisons (2%) inferior. In 20 instances our assessment differed from the authors' verdict: in 9 instances we judged the result to be superior where the authors' verdict was non-inferior, in 6 instances to be inconclusive as opposed to inferior and once to be non-inferior instead of inferior. The authors' verdict was more favourable to the experimental treatment in four comparisons, each time judging the result to be non-inferior instead of inconclusive.

Meta-analysis

The funnel plot showed a symmetrical distribution of results around RR 1 (Figure 3). Forty-seven percent of comparisons ($n = 82$) had a point estimate >1 (favouring standard treatment) and 53% ($n = 93$) <1 (favouring experimental treatment). Of the 130 comparisons judged to be non-inferior, the point estimate favoured experimental treatment in 58% ($n = 76$) and standard treatment in 42% ($n = 54$). The combined RR for all 175 comparisons was 0.994 (95% CI 0.978–1.010) using a random-effects model and 1.002 (95% CI 0.966–1.008) using a fixed-effects model. The combined RR for comparisons judged to be non-inferior was 0.995 (95% CI 0.983–1.006). Table 2 shows stratified random-effects meta-analyses according to trial result, measure of effect, type of endpoint, source of funding, by two journal strata and according to whether the judgement of the result was based on an intention-to-treat analysis or not. Using a random-effects model, the combined estimate for trials funded by industry was 0.978 (95% CI 0.956–1.000). The combined result for trials funded by public sources was 1.008 (95% CI 0.980–1.038). These two estimates did not differ significantly ($P = 0.15$ from random-effects meta-regression). All meta-analyses were also performed using a fixed-effects model and are presented in Appendix Table 4 available as supplementary data at *IJE* online. The main result and the results from the stratified analyses were similar for the random- and fixed-effects meta-analyses except for a difference in the result stratified by funding source.

Non-inferiority margins

The margin was expressed as an RR for 33 comparisons and could be converted from a risk difference to a RR for 91 comparisons. The median pre-specified non-inferiority margin was 1.31 [inter-quartile range (IQR) 1.18–1.59]. Stratified according to trial result, the median margin was 1.42 (range 1.21–4.75) for 3 comparisons judged to be inferior, 1.33 (IQR 1.14–1.51) for 23 comparisons judged to be inconclusive, 1.31 (IQR 1.19–1.59) for 92 comparisons judged to be non-inferior and 1.45 (IQR 1.20–1.75) for 6 comparisons judged to be superior. Stratified by type of endpoint the median margin was 1.34 (IQR 1.19–1.50) for comparisons that had mortality as (part of a combined) endpoint, 1.38 (IQR 1.20–1.70) for comparisons in which clinical disease was the endpoint

Table 1 Characteristics of the 170 non-inferiority trials included in the meta-analysis

Study characteristics	*n*	Percentage
Journal		
New England Journal of Medicine	33	19
Lancet	28	16
Circulation	11	6
Journal of the American Medical Association	10	6
British Medical Journal	8	5
Obstetrics and Gynaecology	8	5
Paediatrics	8	5
Other (from 36 different journals)	64	38
Field in medicine		
Cardiovascular medicine	47	28
Infectious disease	43	25
Obstetrics and gynaecology	12	7
Oncology	10	6
Rheumatology	10	6
Surgery	9	5
Psychiatry	5	3
Other	34	20
Type of comparison		
Drug	133	78
Procedure	23	14
Device	14	8
Source of funding		
Industry	94	55
Public or charity	46	27
Mixed	19	11
Not reported	11	7
Median number of participants per comparison (range)[a]	467 (40–20 332)	
Effect measure[a]		
Risk difference	106	61
Ratio measure	33	19
HR	20	
Risk ratio	7	
OR	6	
Continuous	36	21
Main analysis used to judge the result[a]		
Intention-to-treat analysis	95	54
Modified intention-to-treat or per protocol analysis	80	46

[a]Characteristics for 175 comparisons from the 170 trials.

and 1.23 (IQR 1.15–1.35) for comparisons reported as a surrogate endpoint. Stratified according to source of funding the median margin was 1.30 (IQR 1.17–1.50) for 62 trials funded by industry, 1.31 (IQR 1.12–1.63) for 33 trials funded by public sources and 1.35 (IQR 1.23–1.55) for 19 trials with mixed funding.

The ratio of the observed and expected incidence of the primary endpoint in the group that received standard treatment could be calculated for 112 comparisons. Fifty-three percent of comparisons ($n = 59$) had a ratio <1, indicating that the standard treatment performed better than was expected at the design stage of the trial and 46% ($n = 51$) >1, indicating that the standard treatment performed worse than was expected. Two ratios were exactly 1. The mean ratio was 0.941 (95% CI 0.859–1.030), meaning that on average the chosen standard treatments performed slightly better than was estimated at the design stage of the trials. Stratified by source of funding, this ratio was 0.974 (95% CI 0.865–1.097) for 55 studies funded by industry and 0.906 (95% CI 0.760–1.080) for 31 studies funded by public sources. These two estimates did not differ significantly ($P = 0.5$ from t-test for equality of means).

Discussion

In this meta-analysis of trials using a non-inferiority design, experimental treatments were regarded as non-inferior to standard treatments in the majority of studies. The combined RR for these studies comparing experimental with standard treatments was close to 1. For non-inferiority trials published in core clinical journals, this finding contradicts the hypothesis that new treatments that gain a verdict of non-inferiority are systematically less effective than standard treatments.

Our study has several strengths and limitations. We aimed to include all the non-inferiority trials published in these journals, irrespective of the type of endpoints or measures of effect. We restricted the search to the group of core clinical journals, as defined in PubMed, which is the same group of journals as in the Abridged Index Medicus (AIM). This selection covers a wide range of journals from many clinical specialties. Our results may therefore be representative for non-inferiority trials published in other journals. However, external validity may be limited to higher quality journals. If non-AIM journals are of lower quality, the characteristics of non-inferiority trials published in those journals might be different. Furthermore, our search would have missed trials that do not mention the non-inferiority design in the abstract, the title or as a key word. The characteristics of such trials might also differ.

We examined two aspects of non-inferiority trials: first, we combined the results of a large number of non-inferiority trials in a meta-analysis; secondly, we examined the non-inferiority margins chosen by the

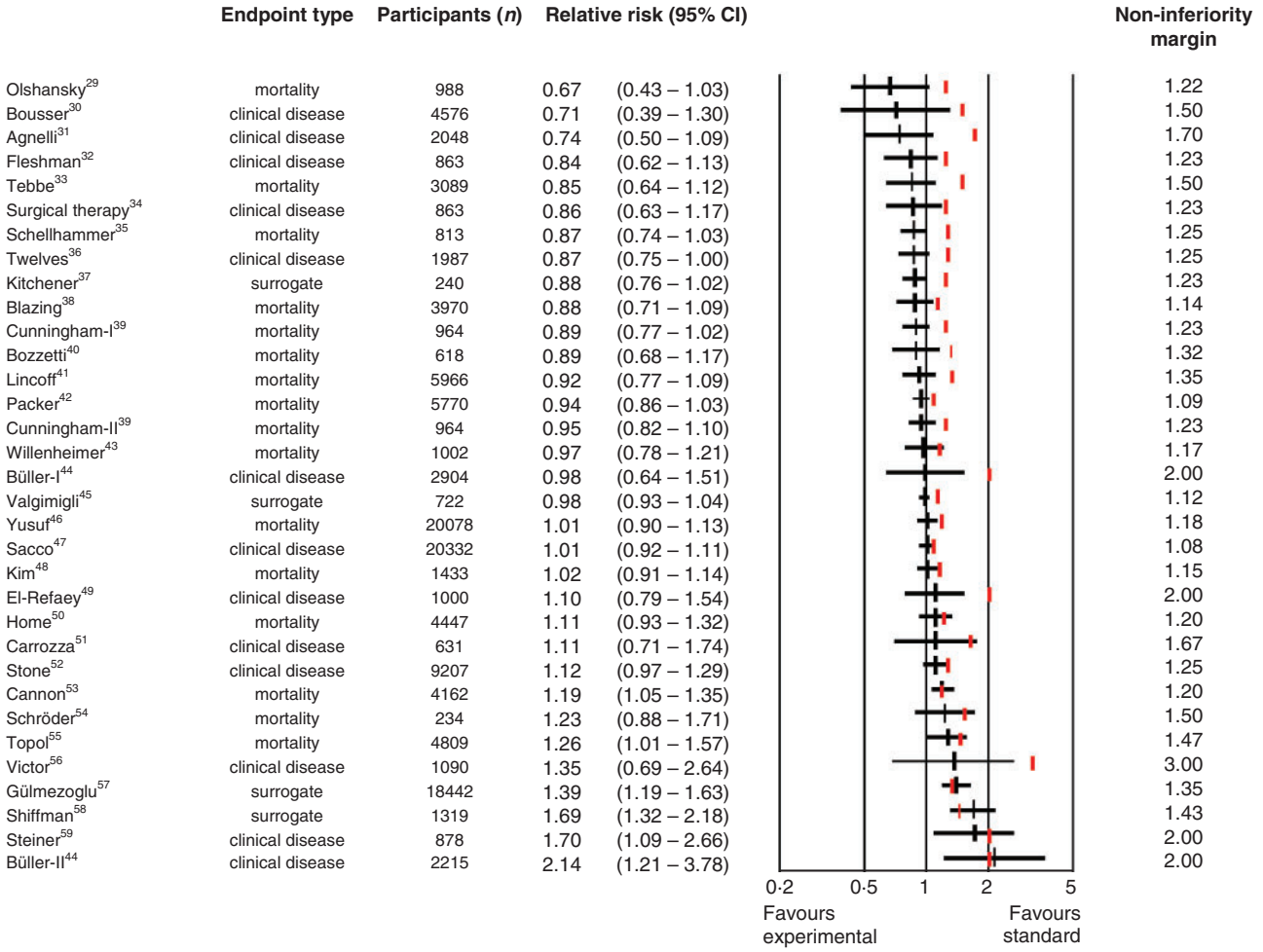

Figure 2 Results for 33 comparisons from 31 trials in which the result was expressed as an RR. Point estimates, CIs and non-inferiority margins (red lines, lighter gray in printed version) are shown

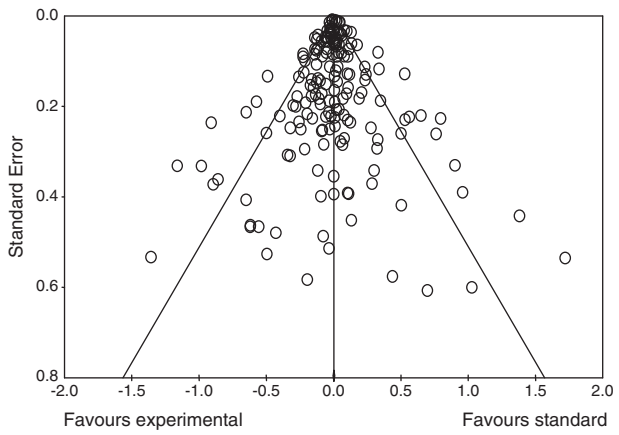

Figure 3 Funnel plot of the standard error by the log RR for 175 comparisons. Treatment on the *X*-axis and standard error on the *Y*-axis. Bias would lead to an asymmetrical appearance of the funnel plot

investigators. Importantly, the combined estimate from the meta-analysis will not be influenced by the choice of the non-inferiority margin. The combined estimate will be influenced by the efficaciousness of standard treatment. If the standard treatment is not effective, the experimental treatment is in fact tested against 'placebo' in a non-inferiority design. Although we did not assess whether the chosen standard treatment represented the best-available comparator, we did assess how standard treatments performed in view of what trialists had expected. On average, the standard treatments performed slightly better than was estimated at the design stage of the trials. Our study did not address other important issues pertaining to non-inferiority trials. For example, we did not assess whether a non-inferiority trial was the appropriate design to use (or whether a superiority design would have been more appropriate) or whether the choice of the non-inferiority margin that was used for the power calculation and statistical testing

Table 2 Random-effects meta-analyses of 175 comparisons of experimental and standard treatments from 170 non-inferiority trials

	Number of comparisons	RR (95% CI)
Overall analysis	175	0.994 (0.978–1.010)
Stratified analyses		
By result		
Result judged as inferior	3	2.255 (1.587–3.204)
Result judged as inconclusive	27	1.163 (1.102–1.227)
Result judged as non-inferior	130	0.995 (0.983–1.006)
Result judged as superior	15	0.694 (0.617–0.780)
By effect measure		
Risk difference	106	0.996 (0.982–1.010)
Ratio measure	33	1.012 (0.958–1.069)
Continuous	36	0.954 (0.826–1.101)
By type of endpoint		
Mortality as (part of a combined) endpoint	35	0.974 (0.935–1.015)
Clinical disease	81	0.998 (0.980–1.017)
Surrogate endpoint	59	1.000 (0.965–1.037)
By source of funding		
Industry	96	0.978 (0.956–1.000)
Public source	48	1.008 (0.980–1.038)
Mixed	20	1.035 (0.972–1.103)
Not reported	11	1.018 (0.930–1.113)
By journal		
N Engl J Med/Lancet/JAMA/BMJ/Ann Intern Med	87	0.990 (0.968–1.012)
Other journals	88	0.999 (0.976–1.023)
By the analysis used to judge the result		
Intention-to-treat analysis	95	1.002 (0.977–1.028)
Modified intention-to-treat- or per-protocol analysis	80	0.989 (0.969–1.009)

All meta-analyses mentioned here were performed using a random-effects model. All meta-analyses were also performed using a fixed-effects model and are shown in Appendix Table 4 available as supplementary data at *IJE* online. The main result and the results from the stratified analyses were similar for the random- and fixed-effects meta-analyses except for a difference in the result stratified by funding source.

made clinical sense. The non-inferiority margin is often criticized as being arbitrary, unacceptably wide or even fraudulent.[8,10] The selection of the non-inferiority margin should be based on a combination of statistical reasoning and clinical judgement.[9,24] Others have reviewed the rationale for the size of the margins in non-inferiority trials.[7,8] They found that the majority of trials did not justify the choice of the margin and that <20% reported a clinical consideration. An in-depth analysis of each trial with subject-matter knowledge on each topic would have been required to judge whether the choice of the margin was adequate. This was beyond the aim of the present analysis.

Does our meta-analysis rebuke some of the criticism aimed at non-inferiority trials? We found that the combined RR for all studies was close to 1. This contradicts the hypothesis that in non-inferiority trials the experimental treatment is generally less effective than the standard treatment. We believe that this is an important, reassuring finding, considering the criticism that has been levelled at non-inferiority trials.[1–12] Several issues should nevertheless be considered when interpreting this result. First, current standards for drug approval stipulate that a new treatment should be better than placebo and (at least) non-inferior to the established options. This means that demonstrating non-inferiority can legally suffice for the licensing of a new drug. The underlying assumption is often that a 'non-inferior' treatment has added value regarding other properties, such as ease of use, lower costs or fewer adverse effects, which might offset a small loss in efficacy. Sometimes such superior properties, such as costs, are self-evident and do not have to be demonstrated in a trial. Claiming that an agent has less adverse

effects should however be based on evidence. A separate analysis of the adverse event data, analyses of combined endpoints or a meta-analysis of several trials might be appropriate and informative to demonstrate superiority in this respect. Of the 175 comparisons in this meta-analysis, we considered the experimental treatment to be non-inferior in 130 (74%) and superior in 15 (9%). Although in 6 of these 145 comparisons the authors deemed the experimental treatment to be clinically inferior based on a secondary endpoint, the majority of published non-inferiority trials can be used to support the registration of a new treatment. The added value and safety of these treatments may not always be self-evident and may not always be demonstrated in the trial. The follow-up time and the sample size of trials are limited, making it improbable that rarer side effects or long-term side effects are detected.

Secondly, for superiority trials, it has repeatedly been described how the outcome may be skewed in favour of the experimental treatment by making convenient choices when designing the study or reporting or publishing the result. This may involve the choice of (the dosage of) the comparator drug, the choice of patients, endpoints or of the type of analysis.[19,27] It may also involve selective reporting of data or changing the pre-specified endpoint after a study is completed.[16] It is plausible that such mechanisms affect the results of non-inferiority trials. In other words, biased choices in study design and bias due to selective reporting of outcomes may make it more likely that an experimental treatment is considered non-inferior after completion of the trial. We did not have access to the study protocols of the included articles and relied on what was reported as the primary endpoint. Also, we restricted our search to studies that have been published. Unpublished trials are more likely to favour standard treatment.[17] Therefore, if publication bias would be an issue, our results might be skewed in favour of experimental treatments. All these potential sources of bias would remain unnoticed in our meta-analysis. Although the funnel plot showed a symmetrical distribution of results around RR 1, this does not rule out biases. This leaves the possibility that our finding of an overall RR close to unity is skewed in favour of experimental treatment. The finding that studies sponsored by industry were more likely to have results favouring sponsored treatments is in line with other reports.[15,25,26] Systematic bias has been suggested as a possible explanation. Our finding could also be due to the play of chance.

Thirdly, the statistical verdict of non-inferiority permits licensing of a drug even if the trial result shows that it is somewhat less effective than standard. Therefore, some treatments that are approved based on non-inferiority testing may be less effective compared with the standard therapy with respect to the primary endpoint. A cascade of non-inferiority trials is possible, in which each next experimental treatment is slightly less effective than the previously established 'standard'. After several generations of non-inferiority trials, ineffective interventions could be licensed, leading to deteriorating patient care.[4,11] This outcome has been called 'bio-creep'.[28] Our results are relevant in this context. Our study showed that of the 130 comparisons judged to be non-inferior, the point estimate favoured the standard treatment in 42% of trials. Biocreep could occur if two or three trials in succession belong to this 42% category and if each next trial adopts the previously demonstrated non-inferior treatment as the new active control treatment. Importantly, our study provided no empirical evidence for or against the existence of biocreep.

In conclusion, the number of non-inferiority trials published in clinical journals has greatly increased. We found that the experimental treatments that gain a verdict of non-inferiority in trials published in core clinical journals are not systematically less effective than the standard treatments. Biases in design, reporting and publication cannot be ruled out and may have skewed the study results in favour of experimental treatments. Continued vigilance is required to assure that non-inferiority trials are used appropriately.

Supplementary data

Supplementary data are available at *IJE* online.

Acknowledgement

We acknowledge the contribution of Theo Stijnen, PhD, Professor of Medical Statistics, Department of Medical Statistics and Bioinformatics, Leiden University Medical Centre, for critical discussion of the analyses, and of J.W. Schoones, MA, Walaeus Library, Leiden University Medical Centre, for assistance with the search strategy.

Conflict of interest: None declared.

KEY MESSAGES

- There is a concern that non-inferiority trials might be deliberately designed to conceal that a new treatment is less effective than a standard treatment. There is little empirical evidence at present to support this notion, however.

- The combined RR from 170 randomized trials using a non-inferiority design and published in core clinical journals in recent years was close to 1, favouring neither the experimental nor the standard treatment.
- In the majority of trials, the new treatments were considered to be non-inferior. For these trials the combined RR was also close to 1.
- The experimental treatments that gain a verdict of non-inferiority do not, therefore, appear to be systematically less effective than the standard treatments.
- The evidence from published non-inferiority trials might still be distorted by publication bias, or by a biased choice of standard treatments. Further studies are required to clarify the risk of bias in non-inferiority trials.

References

[1] Ware JH, Antman EM. Equivalence trials. *N Engl J Med* 1997;**337:**1159–61.

[2] Siegel JP. Equivalence and noninferiority trials. *Am Heart J* 2000;**139:**S166–70.

[3] Djulbegovic B, Clarke M. Scientific and ethical issues in equivalence trials. *JAMA* 2001;**285:**1206–8.

[4] D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med* 2003; **22:**169–86.

[5] James Hung HM, Wang SJ, Tsong Y, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. *Stat Med* 2003;**22:** 213–25.

[6] Kaul S, Diamond GA, Weintraub WS. Trials and tribulations of non-inferiority: the ximelagatran experience. *J Am Coll Cardiol* 2005;**46:**1986–95.

[7] Lange S, Freitag G. Choice of delta: requirements and reality – results of a systematic review. *Biom J* 2005;**47:** 12–27.

[8] Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA* 2006;**295:**1147–51.

[9] Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;**295:**1152–60.

[10] Gotzsche PC. Lessons from and cautions about noninferiority and equivalence randomized trials. *JAMA* 2006;**295:** 1172–74.

[11] Fueglistaler P, Adamina M, Guller U. Non-inferiority trials in surgical oncology. *Ann Surg Oncol* 2007;**14:** 1532–39.

[12] Garattini S, Bertele' V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet* 2007; **370:**1875–77.

[13] Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 2006;**145:**62–69.

[14] Smith R. Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Med* 2005; **2:**e138.

[15] Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;**326:** 1167–70.

[16] Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;**291:**2457–65.

[17] Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;**358:**252–60.

[18] Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;**302:** 977–84.

[19] Leucht S, Heres S, Hamann J, Kissling W. Pretrial medication bias in randomized antipsychotic drug trials. *Am J Psychiatry* 2007;**164:**1266–67.

[20] Dunnett CW, Gent M. An alternative to the use of two-sided tests in clinical trials. *Stat Med* 1996;**15:**1729–38.

[21] The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). E9: statistical principles for clinical trials. *Federal Register* 1998;**63:**49583–98.

[22] Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull* 1995;**117:**167–78.

[23] Carlier C, Coste J, Etchepare M, Periquet B, Amedee-Manesme O. A randomised controlled trial to test equivalence between retinyl palmitate and beta carotene for vitamin A deficiency. *BMJ* 1993;**307:**1106–10.

[24] Committee for Medicinal Products for Human Use (CHMP) guideline on the choice of the non-inferiority margin. *Stat Med* 2006;**25:**1628–38.

[25] Djulbegovic B, Lacevic M, Cantor A *et al*. The uncertainty principle and industry-sponsored research. *Lancet* 2000; **356:**635–38.

[26] Ridker PM, Torres J. Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000–2005. *JAMA* 2006;**295:** 2270–74.

[27] Smith R. Medical journals are an extension of the marketing arm of pharmaceutical companies. *PLoS Med* 2005; **2:**e138.

[28] Fleming TR. Current issues in non-inferiority trials. *Stat Med* 2008;**27:**317–32.

[29] Olshansky B, Day JD, Moore S *et al*. Is dual-chamber programming inferior to single-chamber programming in an implantable cardioverter-defibrillator? Results of the INTRINSIC RV (Inhibition of Unnecessary RV Pacing With AVSH in ICDs) study. *Circulation* 2007;**115:** 9–16.

[30] Bousser MG, Bouthier J, Buller HR *et al*. Comparison of idraparinux with vitamin K antagonists for prevention of thromboembolism in patients with atrial fibrillation: a randomised, open-label, non-inferiority trial. *Lancet* 2008;**371:**315–21.

[31] Agnelli G, Bergqvist D, Cohen AT, Gallus AS, Gent M. Randomized clinical trial of postoperative fondaparinux versus perioperative dalteparin for prevention of venous thromboembolism in high-risk abdominal surgery. *Br J Surg* 2005;**92:**1212–20.

[32] Fleshman J, Sargent DJ, Green E *et al*. Laparoscopic colectomy for cancer is not inferior to open surgery based on 5-year data from the COST Study Group trial. *Ann Surg* 2007;**246:**655–62.

[33] Tebbe U, Michels R, Adgey J *et al*. Randomized, double-blind study comparing saruplase with streptokinase therapy in acute myocardial infarction: the COMPASS Equivalence Trial. Comparison Trial of Saruplase and Streptokinase (COMASS) Investigators. *J Am Coll Cardiol* 1998;**31:**487–93.

[34] A comparison of laparoscopically assisted and open colectomy for colon cancer. *N Engl J Med* 2004;**350:** 2050–59.

[35] Schellhammer PF, Sharifi R, Block NL *et al*. A controlled trial of bicalutamide versus flutamide, each in combination with luteinizing hormone-releasing hormone analogue therapy, in patients with advanced prostate carcinoma. Analysis of time to progression. CASODEX Combination Study Group. *Cancer* 1996;**78:** 2164–69.

[36] Twelves C, Wong A, Nowacki MP *et al*. Capecitabine as adjuvant treatment for stage III colon cancer. *N Engl J Med* 2005;**352:**2696–704.

[37] Kitchener HC, Dunn G, Lawton V *et al*. Laparoscopic versus open colposuspension – results of a prospective randomised controlled trial. *BJOG* 2006;**113:**1007–13.

[38] Blazing MA, de Lemos JA, White HD *et al*. Safety and efficacy of enoxaparin vs unfractionated heparin in patients with non-ST-segment elevation acute coronary syndromes who receive tirofiban and aspirin: a randomized controlled trial. *JAMA* 2004;**292:**55–64.

[39] Cunningham D, Starling N, Rao S *et al*. Capecitabine and oxaliplatin for advanced esophagogastric cancer. *N Engl J Med* 2008;**358:**36–46.

[40] Bozzetti F, Marubini E, Bonfanti G *et al*. Subtotal versus total gastrectomy for gastric cancer: five-year survival rates in a multicenter randomized Italian trial. Italian Gastrointestinal Tumor Study Group. *Ann Surg* 1999; **230:**170–78.

[41] Lincoff AM, Bittl JA, Harrington RA *et al*. Bivalirudin and provisional glycoprotein IIb/IIIa blockade compared with heparin and planned glycoprotein IIb/IIIa blockade during percutaneous coronary intervention: REPLACE-2 randomized trial. *JAMA* 2003;**289:**853–63.

[42] Packer M, Califf RM, Konstam MA *et al*. Comparison of omapatrilat and enalapril in patients with chronic heart failure: the Omapatrilat Versus Enalapril Randomized Trial of Utility in Reducing Events (OVERTURE). *Circulation* 2002;**106:**920–26.

[43] Willenheimer R, van Veldhuisen DJ, Silke B *et al*. Effect on survival and hospitalization of initiating treatment for chronic heart failure with bisoprolol followed by enalapril, as compared with the opposite sequence: results of the randomized Cardiac Insufficiency Bisoprolol Study (CIBIS) III. *Circulation* 2005;**112:**2426–35.

[44] Buller HR, Cohen AT, Davidson B *et al*. Idraparinux versus standard therapy for venous thromboembolic disease. *N Engl J Med* 2007;**357:**1094–104.

[45] Valgimigli M, Campo G, Percoco G *et al*. Comparison of angioplasty with infusion of tirofiban or abciximab and with implantation of sirolimus-eluting or uncoated stents for acute myocardial infarction: the MULTISTRATEGY randomized trial. *JAMA* 2008;**299:**1788–99.

[46] Yusuf S, Mehta SR, Chrolavicius S *et al*. Comparison of fondaparinux and enoxaparin in acute coronary syndromes. *N Engl J Med* 2006;**354:**1464–76.

[47] Sacco RL, Diener HC, Yusuf S *et al*. Aspirin and extended-release dipyridamole versus clopidogrel for recurrent stroke. *N Engl J Med* 2008;**359:**1238–51.

[48] Kim ES, Hirsh V, Mok T *et al*. Gefitinib versus docetaxel in previously treated non-small-cell lung cancer (INTEREST): a randomised phase III trial. *Lancet* 2008; **372:**1809–18.

[49] El-Refaey H, Nooh R, O'Brien P *et al*. The misoprostol third stage of labour study: a randomised controlled comparison between orally administered misoprostol and standard management. *BJOG* 2000;**107:**1104–10.

[50] Home PD, Pocock SJ, Beck-Nielsen H *et al*. Rosiglitazone evaluated for cardiovascular outcomes – an interim analysis. *N Engl J Med* 2007;**357:**28–38.

[51] Carrozza JP Jr, Mumma M, Breall JA *et al*. Randomized evaluation of the TriActiv balloon-protection flush and extraction system for the treatment of saphenous vein graft disease. *J Am Coll Cardiol* 2005;**46:**1677–83.

[52] Stone GW, Bertrand ME, Moses JW *et al*. Routine upstream initiation vs deferred selective use of glycoprotein IIb/IIIa inhibitors in acute coronary syndromes: the ACUITY Timing trial. *JAMA* 2007;**297:**591–602.

[53] Cannon CP, Braunwald E, McCabe CH *et al*. Intensive versus moderate lipid lowering with statins after acute coronary syndromes. *N Engl J Med* 2004;**350:**1495–504.

[54] Schroder FH, Kurth KH, Fossa SD *et al*. Early versus delayed endocrine treatment of pN1–3 M0 prostate cancer without local treatment of the primary tumor: results of European Organisation for the Research and Treatment of Cancer 30846 – a phase III study. *J Urol* 2004;**172:**923–27.

[55] Topol EJ, Moliterno DJ, Herrmann HC *et al*. Comparison of two platelet glycoprotein IIb/IIIa inhibitors, tirofiban and abciximab, for the prevention of ischemic events with percutaneous coronary revascularization. *N Engl J Med* 2001;**344:**1888–94.

[56] Victor JC, Monto AS, Surdina TY *et al*. Hepatitis A vaccine versus immune globulin for postexposure prophylaxis. *N Engl J Med* 2007;**357:**1685–94.

[57] Gulmezoglu AM, Villar J, Ngoc NT *et al*. WHO multicentre randomised trial of misoprostol in the management of the third stage of labour. *Lancet* 2001;**358:**689–95.

[58] Shiffman ML, Suter F, Bacon BR *et al*. Peginterferon alfa-2a and ribavirin for 16 or 24 weeks in HCV genotype 2 or 3. *N Engl J Med* 2007;**357:**124–34.

[59] Steiner MJ, Dominik R, Rountree RW, Nanda K, Dorflinger LJ. Contraceptive effectiveness of a polyurethane condom and a latex condom: a randomized controlled trial. *Obstet Gynecol* 2003;**101:**539–47.

[60] Lai CL, Gane E, Liaw YF *et al*. Telbivudine versus lamivudine in patients with chronic hepatitis B. *N Engl J Med* 2007;**357:**2576–88.

[61] Madruga JV, Berger D, McMurchie M *et al*. Efficacy and safety of darunavir-ritonavir compared with that of lopinavir-ritonavir at 48 weeks in treatment-experienced, HIV-infected patients in TITAN: a randomised controlled phase III trial. *Lancet* 2007;**370:**49–58.

[62] Reboli AC, Rotstein C, Pappas PG *et al*. Anidulafungin versus fluconazole for invasive candidiasis. *N Engl J Med* 2007;**356:**2472–82.

[63] Zongo I, Dorsey G, Rouamba N *et al*. Artemether-lumefantrine versus amodiaquine plus sulfadoxine-pyrimethamine for uncomplicated falciparum malaria in Burkina Faso: a randomised non-inferiority trial. *Lancet* 2007;**369:**491–98.

[64] Gallant JE, DeJesus E, Arribas JR *et al*. Tenofovir DF, emtricitabine, and efavirenz vs. zidovudine, lamivudine, and efavirenz for HIV. *N Engl J Med* 2006;**354:**251–60.

[65] Ginzler EM, Dooley MA, Aranow C *et al*. Mycophenolate mofetil or intravenous cyclophosphamide for lupus nephritis. *N Engl J Med* 2005;**353:**2219–28.

[66] Bernard P, Chosidow O, Vaillant L. Oral pristinamycin versus standard penicillin regimen to treat erysipelas in adults: randomised, non-inferiority, open trial. *BMJ* 2002; **325:**864.

[67] Zepp F, Schuind A, Meyer C *et al*. Safety and reactogenicity of a novel DTPa-HBV-IPV combined vaccine given along with commercial Hib vaccines in comparison with separate concomitant administration of DTPa, Hib, and OPV vaccines in infants. *Pediatrics* 2002;**109:**e58.

[68] Danel C, Moh R, Chaix ML *et al*. Two-months-off, four-months-on antiretroviral regimen increases the risk of resistance, compared with continuous therapy: a randomized trial involving West African adults. *J Infect Dis* 2009;**199:**66–76.

[69] Kruis W, Kiudelis G, Racz I *et al*. Once daily versus three times daily mesalazine granules in active ulcerative colitis: a double-blind, double-dummy, randomised, non-inferiority trial. *Gut* 2009;**58:**233–40.

[70] Abdulla S, Sagara I, Borrmann S *et al*. Efficacy and safety of artemether-lumefantrine dispersible tablets compared with crushed commercial tablets in African infants and children with uncomplicated malaria: a randomised, single-blind, multicentre trial. *Lancet* 2008;**372:**1819–27.

[71] Marano N, Plikaytis BD, Martin SW *et al*. Effects of a reduced dose schedule and intramuscular administration of anthrax vaccine adsorbed on immunogenicity and safety at 7 months: a randomized trial. *JAMA* 2008;**300:** 1532–43.

[72] Windecker S, Serruys PW, Wandel S *et al*. Biolimus-eluting stent with biodegradable polymer versus sirolimus-eluting stent with durable polymer for coronary revascularisation (LEADERS): a randomised non-inferiority trial. *Lancet* 2008;**372:**1163–73.

[73] Molina JM, Andrade-Villanueva J, Echevarria J *et al*. Once-daily atazanavir/ritonavir versus twice-daily lopinavir/ritonavir, each in combination with tenofovir and emtricitabine, for management of antiretroviral-naive HIV-1-infected patients: 48 week efficacy and safety results of the CASTLE study. *Lancet* 2008;**372:**646–55.

[74] Engler RJ, Nelson MR, Klote MM *et al*. Half- vs full-dose trivalent inactivated influenza vaccine (2004–2005): age, dose, and sex effects on immune responses. *Arch Intern Med* 2008;**168:**2405–14.

[75] Montini G, Rigon L, Zucchetta P *et al*. Prophylaxis after first febrile urinary tract infection in children? A multicenter, randomized, controlled, noninferiority trial. *Pediatrics* 2008;**122:**1064–71.

[76] Stabile E, Nammas W, Salemme L *et al*. The CIAO (Coronary Interventions Antiplatelet-based Only) Study: a randomized study comparing standard anticoagulation regimen to absence of anticoagulation for elective percutaneous coronary intervention. *J Am Coll Cardiol* 2008;**52:** 1293–98.

[77] Lennon DR, Farrell E, Martin DR, Stewart JM. Once-daily amoxicillin versus twice-daily penicillin V in group A beta-haemolytic streptococcal pharyngitis. *Arch Dis Child* 2008;**93:**474–78.

[78] Barber MD, Kleeman S, Karram MM *et al*. Transobturator tape compared with tension-free vaginal tape for the treatment of stress urinary incontinence: a randomized controlled trial. *Obstet Gynecol* 2008;**111:**611–21.

[79] Contant CM, Hop WC, van't Sant HP *et al*. Mechanical bowel preparation for elective colorectal surgery: a multicentre randomised trial. *Lancet* 2007;**370:**2112–17.

[80] Tauber E, Kollaritsch H, Korinek M *et al*. Safety and immunogenicity of a Vero-cell-derived, inactivated Japanese encephalitis vaccine: a non-inferiority, phase III, randomised controlled trial. *Lancet* 2007;**370:**1847–53.

[81] Eriksson BI, Dahl OE, Rosencher N *et al*. Dabigatran etexilate versus enoxaparin for prevention of venous thromboembolism after total hip replacement: a randomised, double-blind, non-inferiority trial. *Lancet* 2007; **370:**949–56.

[82] Montini G, Toffolo A, Zucchetta P *et al*. Antibiotic treatment for pyelonephritis in children: multicentre randomised controlled non-inferiority trial. *BMJ* 2007;**335:**386.

[83] Sundar S, Jha TK, Thakur CP, Sinha PK, Bhattacharya SK. Injectable paromomycin for Visceral leishmaniasis in India. *N Engl J Med* 2007;**356:**2571–81.

[84] von Hertzen H, Piaggio G, Huong NT *et al*. Efficacy of two intervals and two routes of administration of misoprostol for termination of early pregnancy: a randomised controlled equivalence trial. *Lancet* 2007;**369:**1938–46.

[85] Kuse ER, Chetchotisakd P, da Cunha CA *et al*. Micafungin versus liposomal amphotericin B for candidaemia and invasive candidosis: a phase III randomised double-blind trial. *Lancet* 2007;**369:**1519–27.

[86] Lacroix J, Hebert PC, Hutchison JS *et al*. Transfusion strategies for patients in pediatric intensive care units. *N Engl J Med* 2007;**356:**1609–19.

[87] Heijnen EM, Eijkemans MJ, De Klerk C *et al*. A mild treatment strategy for in-vitro fertilisation: a randomised non-inferiority trial. *Lancet* 2007;**369:**743–49.

[88] Mauri L, Cox D, Hermiller J *et al*. The PROXIMAL trial: proximal protection during saphenous vein graft intervention using the Proxis Embolic Protection System: a randomized, prospective, multicenter clinical trial. *J Am Coll Cardiol* 2007;**50:**1442–49.

[89] Heidegger T, Starzyk L, Villiger CR *et al*. Fiberoptic intubation and laryngeal morbidity: a randomized controlled trial. *Anesthesiology* 2007;**107:**585–90.

[90] Barnhart KT, Rosenberg MJ, MacKay HT *et al*. Contraceptive efficacy of a novel spermicidal microbicide used

with a diaphragm: a randomized controlled trial. *Obstet Gynecol* 2007;**110:**577–86.

[91] McEvoy JP, Lieberman JA, Perkins DO *et al*. Efficacy and tolerability of olanzapine, quetiapine, and risperidone in the treatment of early psychosis: a randomized, double-blind 52-week comparison. *Am J Psychiatry* 2007;**164:** 1050–60.

[92] Jones TK, Latson LA, Zahn E *et al*. Results of the U.S. multicenter pivotal study of the HELEX septal occluder for percutaneous closure of secundum atrial septal defects. *J Am Coll Cardiol* 2007;**49:**2215–21.

[93] Heyde GS, Koch KT, de Winter RJ *et al*. Randomized trial comparing same-day discharge with overnight hospital stay after percutaneous coronary intervention: results of the Elective PCI in Outpatient Study (EPOS). *Circulation* 2007;**115:**2299–306.

[94] Creinin MD, Schreiber CA, Bednarek P *et al*. Mifepristone and misoprostol administered simultaneously versus 24 hours apart for abortion: a randomized controlled trial. *Obstet Gynecol* 2007;**109:**885–94.

[95] Brodie MJ, Perucca E, Ryvlin P, Ben-Menachem E, Meencke HJ. Comparison of levetiracetam and controlled-release carbamazepine in newly diagnosed epilepsy. *Neurology* 2007;**68:**402–8.

[96] Kearon C, Ginsberg JS, Julian JA *et al*. Comparison of fixed-dose weight-adjusted unfractionated heparin and low-molecular-weight heparin for acute treatment of venous thromboembolism. *JAMA* 2006;**296:**935–42.

[97] Fowler VG Jr, Boucher HW, Corey GR *et al*. Daptomycin versus standard therapy for bacteremia and endocarditis caused by Staphylococcus aureus. *N Engl J Med* 2006; **355:**653–65.

[98] Eron J Jr, Yeni P, Gathe J Jr *et al*. The KLEAN study of fosamprenavir-ritonavir versus lopinavir-ritonavir, each in combination with abacavir-lamivudine, for initial treatment of HIV infection over 48 weeks: a randomised non-inferiority trial. *Lancet* 2006;**368:**476–82.

[99] el Moussaoui R, de Borgie CA, van den BP *et al*. Effectiveness of discontinuing antibiotic treatment after three days versus eight days in mild to moderate-severe community acquired pneumonia: randomised, double blind study. *BMJ* 2006;**332:**1355.

[100] Heal C, Buettner P, Raasch B *et al*. Can sutures get wet? Prospective randomised controlled trial of wound management in general practice. *BMJ* 2006;**332:**1053–56.

[101] Bertrand OF, De Larochelliere R, Rodes-Cabau J *et al*. A randomized study comparing same-day home discharge and abciximab bolus only to overnight hospitalization and abciximab bolus and infusion after transradial coronary stent implantation. *Circulation* 2006;**114:**2636–43.

[102] Creinin MD, Schlaff W, Archer DF *et al*. Progesterone receptor modulator for emergency contraception: a randomized controlled trial. *Obstet Gynecol* 2006;**108:** 1089–97.

[103] Bajetta E, Procopio G, Catena L *et al*. Lanreotide autogel every 6 weeks compared with Lanreotide microparticles every 3 weeks in patients with well differentiated neuroendocrine tumors: a Phase III Study. *Cancer* 2006;**107:** 2474–81.

[104] Walter EB, Neuzil KM, Zhu Y *et al*. Influenza vaccine immunogenicity in 6- to 23-month-old children: are identical antigens necessary for priming? *Pediatrics* 2006;**118:**e570–78.

[105] van Mastrigt GA, Heijmans J, Severens JL *et al*. Short-stay intensive care after coronary artery bypass surgery: randomized clinical trial on safety and cost-effectiveness. *Crit Care Med* 2006;**34:**65–75.

[106] Petri M, Kim MY, Kalunian KC *et al*. Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med* 2005;**353:**2550–58.

[107] Kullberg BJ, Sobel JD, Ruhnke M *et al*. Voriconazole versus a regimen of amphotericin B followed by fluconazole for candidaemia in non-neutropenic patients: a randomised non-inferiority trial. *Lancet* 2005;**366:** 1435–42.

[108] Riedner G, Rusizoka M, Todd J *et al*. Single-dose azithromycin versus penicillin G benzathine for the treatment of early syphilis. *N Engl J Med* 2005;**353:**1236–44.

[109] Vincenti F, Larsen C, Durrbach A *et al*. Costimulation blockade with belatacept in renal transplantation. *N Engl J Med* 2005;**353:**770–81.

[110] Nathan N, Borel T, Djibo A *et al*. Ceftriaxone as effective as long-acting chloramphenicol in short-course treatment of meningococcal meningitis during epidemics: a randomised non-inferiority study. *Lancet* 2005;**366:** 308–13.

[111] Oliver RT, Mason MD, Mead GM *et al*. Radiotherapy versus single-dose carboplatin in adjuvant treatment of stage I seminoma: a randomised trial. *Lancet* 2005;**366:** 293–300.

[112] Buyon JP, Petri MA, Kim MY *et al*. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. *Ann Intern Med* 2005;**142:** 953–62.

[113] Albers GW, Diener HC, Frison L *et al*. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: a randomized trial. *JAMA* 2005; **293:**690–98.

[114] Fiessinger JN, Huisman MV, Davidson BL *et al*. Ximelagatran vs low-molecular-weight heparin and warfarin for the treatment of deep vein thrombosis: a randomized trial. *JAMA* 2005;**293:**681–89.

[115] Fogarty C, de WR, Mandell L *et al*. Five-day telithromycin once daily is as effective as 10-day clarithromycin twice daily for the treatment of acute exacerbations of chronic bronchitis and is associated with reduced healthcare resource utilization. *Chest* 2005;**128:**1980–88.

[116] De GK, Rasmussen N, Bacon PA *et al*. Randomized trial of cyclophosphamide versus methotrexate for induction of remission in early systemic antineutrophil cytoplasmic antibody-associated vasculitis. *Arthritis Rheum* 2005;**52:** 2461–69.

[117] Leroy O, Saux P, Bedos JP, Caulin E. Comparison of levofloxacin and cefotaxime combined with ofloxacin for ICU patients with community-acquired pneumonia who do not require vasopressors. *Chest* 2005;**128:**172–83.

[118] Tohen M, Greil W, Calabrese JR *et al*. Olanzapine versus lithium in the maintenance treatment of bipolar disorder: a 12-month, randomized, double-blind, controlled clinical trial. *Am J Psychiatry* 2005;**162:**1281–90.

[119] Conrad SA, Gabrielli A, Margolis B *et al*. Randomized, double-blind comparison of immediate-release omeprazole oral suspension versus intravenous cimetidine for the prevention of upper gastrointestinal bleeding in critically ill patients. *Crit Care Med* 2005;**33:**760–65.

[120] Sinha SK, Lacaze-Masmonteil T, Soler A *et al*. A multicenter, randomized, controlled trial of lucinactant versus poractant alfa among very premature infants at high risk for respiratory distress syndrome. *Pediatrics* 2005;**115:** 1030–38.

[121] Yadav JS, Wholey MH, Kuntz RE *et al*. Protected carotid-artery stenting versus endarterectomy in high-risk patients. *N Engl J Med* 2004;**351:**1493–501.

[122] de Kraker J, Graf N, van Tinteren H *et al*. Reduction of postoperative chemotherapy in children with stage I intermediate-risk and anaplastic Wilms' tumour (SIOP 93–01 trial): a randomised controlled trial. *Lancet* 2004; **364:**1229–35.

[123] Walsh TJ, Teppler H, Donowitz GR *et al*. Caspofungin versus liposomal amphotericin B for empirical antifungal therapy in patients with persistent fever and neutropenia. *N Engl J Med* 2004;**351:**1391–402.

[124] Schmid C, Nkunku S, Merolle A, Vounatsou P, Burri C. Efficacy of 10-day melarsoprol schedule 2 years after treatment for late-stage gambiense sleeping sickness. *Lancet* 2004;**364:**789–90.

[125] Kruis W, Fric P, Pokrotnieks J *et al*. Maintaining remission of ulcerative colitis with the probiotic Escherichia coli Nissle 1917 is as effective as with standard mesalazine. *Gut* 2004;**53:**1617–23.

[126] Conte PF, Guarneri V, Bruzzi P *et al*. Concomitant versus sequential administration of epirubicin and paclitaxel as first-line therapy in metastatic breast carcinoma: results for the Gruppo Oncologico Nord Ovest randomized trial. *Cancer* 2004;**101:**704–12.

[127] McCullough J, Vesole DH, Benjamin RJ *et al*. Therapeutic efficacy and safety of platelets treated with a photochemical process for pathogen inactivation: the SPRINT Trial. *Blood* 2004;**104:**1534–41.

[128] Creinin MD, Fox MC, Teal S *et al*. A randomized comparison of misoprostol 6 to 8 hours versus 24 hours after mifepristone for abortion. *Obstet Gynecol* 2004;**103:** 851–59.

[129] Vogel T, Verreault R, Gourdeau M *et al*. Optimal duration of antibiotic therapy for uncomplicated urinary tract infection in older women: a double-blind randomized controlled trial. *CMAJ* 2004;**170:**469–73.

[130] Stellbrink C, Nixdorff U, Hofmann T *et al*. Safety and efficacy of enoxaparin compared with unfractionated heparin and oral anticoagulants for prevention of thromboembolic complications in cardioversion of nonvalvular atrial fibrillation: the Anticoagulation in Cardioversion using Enoxaparin (ACE) trial. *Circulation* 2004;**109:**997–1003.

[131] Buller HR, Davidson BL, Decousus H *et al*. Subcutaneous fondaparinux versus intravenous unfractionated heparin in the initial treatment of pulmonary embolism. *N Engl J Med* 2003;**349:**1695–702.

[132] Lara LF, Cisneros G, Gurney M *et al*. One-day quadruple therapy compared with 7-day triple therapy for Helicobacter pylori infection. *Arch Intern Med* 2003;**163:** 2079–84.

[133] Stone GW, Rogers C, Hermiller J *et al*. Randomized comparison of distal protection with a filter-based catheter and a balloon occlusion and aspiration system during percutaneous intervention of diseased saphenous vein aorto-coronary bypass grafts. *Circulation* 2003;**108:** 548–53.

[134] Kleber FX, Witt C, Vogel G *et al*. Randomized comparison of enoxaparin with unfractionated heparin for the prevention of venous thromboembolism in medical patients with heart failure or severe respiratory disease. *Am Heart J* 2003;**145:**614–21.

[135] Schiele F, Meneveau N, Gilard M *et al*. Intravascular ultrasound-guided balloon angioplasty compared with stent: immediate and 6-month results of the multicenter, randomized Balloon Equivalent to Stent Study (BEST). *Circulation* 2003;**107:**545–51.

[136] Van Gelder I, Hagens VE, Bosker HA *et al*. A comparison of rate control and rhythm control in patients with recurrent persistent atrial fibrillation. *N Engl J Med* 2002;**347:**1834–40.

[137] Brito FS Jr, Caixeta AM, Perin MA *et al*. Comparison of direct stenting versus stenting with predilation for the treatment of selected coronary narrowings. *Am J Cardiol* 2002;**89:**115–20.

[138] Boogaerts M, Winston DJ, Bow EJ *et al*. Intravenous and oral itraconazole versus intravenous amphotericin B deoxycholate as empirical antifungal therapy for persistent fever in neutropenic patients with cancer who are receiving broad-spectrum antibacterial therapy. A randomized, controlled trial. *Ann Intern Med* 2001;**135:**412–22.

[139] Ross AM, Molhoek P, Lundergan C *et al*. Randomized comparison of enoxaparin, a low-molecular-weight heparin, with unfractionated heparin adjunctive to recombinant tissue plasminogen activator thrombolysis and aspirin: second trial of Heparin and Aspirin Reperfusion Therapy (HART II). *Circulation* 2001;**104:** 648–52.

[140] Baim DS, Cutlip DE, Midei M *et al*. Final results of a randomized trial comparing the MULTI-LINK stent with the Palmaz-Schatz stent for narrowings in native coronary arteries. *Am J Cardiol* 2001;**87:**157–62.

[141] Hogh B, Clarke PD, Camus D *et al*. Atovaquone-proguanil versus chloroquine-proguanil for malaria prophylaxis in non-immune travellers: a randomised, double-blind study. Malarone International Study Team. *Lancet* 2000; **356:**1888–94.

[142] Lallemant M, Jourdain G, Le Coeur S *et al*. A trial of shortened zidovudine regimens to prevent mother-to-child transmission of human immunodeficiency virus type 1. Perinatal HIV Prevention Trial (Thailand) Investigators. *N Engl J Med* 2000;**343:**982–91.

[143] Adam D, Scholz H, Helmerking M. Short-course antibiotic treatment of 4782 culture-proven cases of group A streptococcal tonsillopharyngitis and incidence of post-streptococcal sequelae. *J Infect Dis* 2000;**182:**509–16.

[144] Van De Werf F, Adgey J, Ardissino D *et al*. Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: the ASSENT-2 double-blind randomised trial. *Lancet* 1999;**354:**716–22.

[145] Kern WV, Cometta A, De Bock R *et al*. Oral versus intravenous empirical antimicrobial therapy for fever in patients with granulocytopenia who are receiving cancer chemotherapy. International Antimicrobial Therapy Cooperative Group of the European Organization for Research and Treatment of Cancer. *N Engl J Med* 1999;**341:**312–18.

[146] Lang J, Zuckerman J, Clarke P *et al*. Comparison of the immunogenicity and safety of two 17D yellow fever vaccines. *Am J Trop Med Hyg* 1999;**60:**1045–50.

[147] Low-molecular-weight heparin in the treatment of patients with venous thromboembolism. The Columbus Investigators. *N Engl J Med* 1997;**337:**657–62.

[148] Randomised, double-blind comparison of reteplase double-bolus administration with streptokinase in acute myocardial infarction (INJECT): trial to investigate equivalence. International Joint Efficacy Comparison of Thrombolytics. *Lancet* 1995;**346:**329–36.

[149] Lumbiganon P, Villar J, Laopaiboon M *et al*. One-day compared with 7-day nitrofurantoin for asymptomatic bacteriuria in pregnancy: a randomized controlled trial. *Obstet Gynecol* 2009;**113:**339–45.

[150] Calder AA, Loughney AD, Weir CJ, Barber JW. Induction of labour in nulliparous and multiparous women: a UK, multicentre, open-label study of intravaginal misoprostol in comparison with dinoprostone. *BJOG* 2008;**115:** 1279–88.

[151] Ringleb PA, Allenberg J, Bruckmann H *et al*. 30 day results from the SPACE trial of stent-protected angioplasty versus carotid endarterectomy in symptomatic patients: a randomised non-inferiority trial. *Lancet* 2006;**368:**1239–47.

[152] Lemann M, Mary JY, Colombel JF *et al*. A randomized, double-blind, controlled withdrawal trial in Crohn's disease patients in long-term remission on azathioprine. *Gastroenterology* 2005;**128:**1812–18.

[153] Spandorfer PR, Alessandrini EA, Joffe MD, Localio R, Shaw KN. Oral versus intravenous rehydration of moderately dehydrated children: a randomized, controlled trial. *Pediatrics* 2005;**115:**295–301.

[154] Le SN, Gaboury I, Baird M *et al*. A randomized, double--blind, placebo-controlled noninferiority trial of amoxicillin for clinically diagnosed acute otitis media in children 6 months to 5 years of age. *CMAJ* 2005;**172:**335–41.

[155] Gallant JE, Staszewski S, Pozniak AL *et al*. Efficacy and safety of tenofovir DF vs stavudine in combination therapy in antiretroviral-naive patients: a 3-year randomized trial. *JAMA* 2004;**292:**191–201.

[156] Laiskonis A, Thune T, Neldam S, Hiltunen-Back E. Valacyclovir in the treatment of facial herpes simplex virus infection. *J Infect Dis* 2002;**186(Suppl 1):**S66–70.

[157] A comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction. The Continuous Infusion versus Double-Bolus Administration of Alteplase (COBALT) Investigators. *N Engl J Med* 1997;**337:**1124–30.

[158] Girling DJ. Comparison of oral etoposide and standard intravenous multidrug chemotherapy for small-cell lung cancer: a stopped multicentre randomised trial. Medical Research Council Lung Cancer Working Party. *Lancet* 1996;**348:**563–66.

[159] Mas JL, Chatellier G, Beyssen B *et al*. Endarterectomy versus stenting in patients with symptomatic severe carotid stenosis. *N Engl J Med* 2006;**355:**1660–71.

[160] Sparrow A, Geelhoed G. Prednisolone versus dexamethasone in croup: a randomised equivalence trial. *Arch Dis Child* 2006;**91:**580–83.

[161] Dickinson JE, Evans SF. A comparison of oral misoprostol with vaginal misoprostol administration in second-trimester pregnancy termination for fetal abnormality. *Obstet Gynecol* 2003;**101:**1294–99.

[162] Drucker DJ, Buse JB, Taylor K *et al*. Exenatide once weekly versus twice daily for the treatment of type 2 diabetes: a randomised, open-label, non-inferiority study. *Lancet* 2008;**372:**1240–50.

[163] Stone GW, Midei M, Newman W *et al*. Comparison of an everolimus-eluting stent and a paclitaxel-eluting stent in patients with coronary artery disease: a randomized trial. *JAMA* 2008;**299:**1903–13.

[164] Boutis K, Willan AR, Babyn P *et al*. A randomized, controlled trial of a removable brace versus casting in children with low-risk ankle fractures. *Pediatrics* 2007;**119:** e1256–63.

[165] Calderon Y, Haughey M, Bijur PE *et al*. An educational HIV pretest counseling video program for off-hours testing in the emergency department. *Ann Emerg Med* 2006; **48:**21–27.

[166] Delmas PD, Adami S, Strugala C *et al*. Intravenous ibandronate injections in postmenopausal women with osteoporosis: one-year results from the dosing intravenous administration study. *Arthritis Rheum* 2006;**54:** 1838–46.

[167] Tinmouth J, Kandel G, Tomlinson G *et al*. The effect of dairy product ingestion on human immunodeficiency virus-related diarrhea in a sample of predominantly gay men: a randomized, controlled, double-blind, crossover trial. *Arch Intern Med* 2006;**166:**1178–83.

[168] Szegedi A, Kohnen R, Dienel A, Kieser M. Acute treatment of moderate to severe depression with hypericum extract WS 5570 (St John's wort): randomised controlled double blind non-inferiority trial versus paroxetine. *BMJ* 2005;**330:**503.

[169] Solomon SD, Appelbaum E, Manning WJ *et al*. Effect of the direct Renin inhibitor aliskiren, the Angiotensin receptor blocker losartan, or both on left ventricular mass in patients with hypertension and left ventricular hypertrophy. *Circulation* 2009;**119:**530–37.

[170] Bretzel RG, Nuber U, Landgraf W *et al*. Once-daily basal insulin glargine versus thrice-daily prandial insulin lispro in people with type 2 diabetes on oral hypoglycaemic agents (APOLLO): an open randomised controlled trial. *Lancet* 2008;**371:**1073–84.

[171] Dennehy PH, Bertrand HR, Silas PE *et al*. Coadministration of RIX4414 oral human rotavirus vaccine does not impact the immune response to antigens contained in routine infant vaccines in the United States. *Pediatrics* 2008;**122:**e1062–66.

[172] Maltais F, Bourbeau J, Shapiro S *et al*. Effects of home-based pulmonary rehabilitation in patients with chronic obstructive pulmonary disease: a randomized trial. *Ann Intern Med* 2008;**149:**869–78.

[173] Eranti S, Mogg A, Pluck G *et al*. A randomized, controlled trial with 6-month follow-up of repetitive transcranial magnetic stimulation and electroconvulsive therapy for severe depression. *Am J Psychiatry* 2007;**164:**73–81.

[174] Levin NW, Fishbane S, Canedo FV *et al*. Intravenous methoxy polyethylene glycol-epoetin beta for haemoglobin control in patients with chronic kidney disease who are on dialysis: a randomised non-inferiority trial (MAXIMA). *Lancet* 2007;**370:**1415–21.

[175] Willburger RE, Mysler E, Derbot J *et al*. Lumiracoxib 400 mg once daily is comparable to indomethacin 50 mg three times daily for the treatment of acute flares of gout. *Rheumatology (Oxford)* 2007;**46:**1126–32.

[176] Bingham CO III, Sebba AI, Rubin BR *et al*. Efficacy and safety of etoricoxib 30 mg and celecoxib 200 mg in the

treatment of osteoarthritis in two identically designed, randomized, placebo-controlled, non-inferiority studies. *Rheumatology (Oxford)* 2007;**46:**496–507.

[177] Dalton JD Jr, Schweinle JE. Randomized controlled non-inferiority trial to compare extended release acetaminophen and ibuprofen for the treatment of ankle sprains. *Ann Emerg Med* 2006;**48:**615–23.

[178] Gayer S, Denham D, Alarakhia K *et al*. Ocular decompression devices: liquid mercury balloon vs the tungsten powder balloon. *Am J Ophthalmol* 2006;**142:**500–1.

[179] Lovell K, Cox D, Haddock G *et al*. Telephone administered cognitive behaviour therapy for treatment of obsessive compulsive disorder: randomised controlled non-inferiority trial. *BMJ* 2006;**333:**883.

[180] Mehilli J, Kastrati A, Wessely R *et al*. Randomized trial of a nonpolymer-based rapamycin-eluting stent versus a polymer-based paclitaxel-eluting stent for the reduction of late lumen loss. *Circulation* 2006;**113:**273–79.

[181] Puhan MA, Busching G, Schunemann HJ *et al*. Interval versus continuous high-intensity exercise in chronic obstructive pulmonary disease: a randomized trial. *Ann Intern Med* 2006;**145:**816–25.

[182] Roberts N, Boehm M, Bates M *et al*. Two-center prospective randomized controlled trial of Blake versus Portex drains after cardiac surgery. *J Thorac Cardiovasc Surg* 2006; **132:**1042–46.

[183] Benecke R, Jost WH, Kanovsky P *et al*. A new botulinum toxin type A free of complexing proteins for treatment of cervical dystonia. *Neurology* 2005;**64:**1949–51.

[184] Bobat R, Coovadia H, Stephen C *et al*. Safety and efficacy of zinc supplementation for children with HIV-1 infection in South Africa: a randomised double-blind placebo-controlled trial. *Lancet* 2005;**366:**1862–67.

[185] Christenson LJ, Phillips PK, Weaver AL, Otley CC. Primary closure vs second-intention treatment of skin punch biopsy sites: a randomized trial. *Arch Dermatol* 2005;**141:**1093–99.

[186] Garcia Garcia ML, Wahn U, Gilles L *et al*. Montelukast, compared with fluticasone, for control of asthma among 6- to 14-year-old patients with mild asthma: the MOSAIC study. *Pediatrics* 2005;**116:**360–69.

[187] Schnitzer TJ, Kivitz AJ, Lipetz RS, Sanders N, Hee A. Comparison of the COX-inhibiting nitric oxide donator AZD3582 and rofecoxib in treating the signs and symptoms of Osteoarthritis of the knee. *Arthritis Rheum* 2005;**53:**827–37.

[188] Barnett AH, Bain SC, Bouter P *et al*. Angiotensin-receptor blockade versus converting-enzyme inhibition in type 2 diabetes and nephropathy. *N Engl J Med* 2004;**351:** 1952–61.

[189] Poor G, Strand V. Efficacy and safety of leflunomide 10 mg versus 20 mg once daily in patients with active rheumatoid arthritis: multinational double-blind, randomized trial. *Rheumatology (Oxford)* 2004;**43:**744–49.

[190] Cooper C, Emkey RD, McDonald RH *et al*. Efficacy and safety of oral weekly ibandronate in the treatment of postmenopausal osteoporosis. *J Clin Endocrinol Metab* 2003;**88:**4609–15.

[191] Gibofsky A, Williams GW, McKenna F, Fort JG. Comparing the efficacy of cyclooxygenase 2-specific inhibitors in treating osteoarthritis: appropriate trial design considerations and results of a randomized, placebo-controlled trial. *Arthritis Rheum* 2003;**48:**3102–11.

[192] Mattsson LA, Christiansen C, Colau JC *et al*. Clinical equivalence of intranasal and oral 17beta-estradiol for postmenopausal symptoms. *Am J Obstet Gynecol* 2000; **182:**545–52.

[193] Weiser M, Strosser W, Klein P. Homeopathic vs conventional treatment of vertigo: a randomized double-blind controlled clinical study. *Arch Otolaryngol Head Neck Surg* 1998;**124:**879–85.

[194] Mintz PD, Bass NM, Petz LD *et al*. Photochemically treated fresh frozen plasma for transfusion of patients with acquired coagulopathy of liver disease. *Blood* 2006;**107:** 3753–60.

[195] Lim Y, Sia AT, Ocampo CE. Comparison of computer integrated patient controlled epidural analgesia vs. conventional patient controlled epidural analgesia for pain relief in labour. *Anaesthesia* 2006;**61:**339–44.

[196] Dibra A, Kastrati A, Mehilli J *et al*. Paclitaxel-eluting or sirolimus-eluting stents to prevent restenosis in diabetic patients. *N Engl J Med* 2005;**353:**663–70.

[197] Klaber Moffett JA, Jackson DA, Richmond S *et al*. Randomised trial of a brief physiotherapy intervention compared with usual physiotherapy for neck pain patients: outcomes and patients' preference. *BMJ* 2005; **330:**75.